PLING Adaptive Opportunity Checking of Web Service with Recommendation System

S. Subbulakshmi, Rahul Varma U, Sangeeta Nair
Department of Computer Science and Applications
Amrita Vishwa Vidyapeetham, Amritapuri, India
subbhulakshmis@gmail.com, rahulvarmau@am.amrita.edu, sangeeta.nair9@gmail.com

Abstract - In the recent advancement of e-commerce, most of the populace in the world are using the World Wide Web for the retrieval of information and to perform various transactions through internet. This has led to the increase in the complexity of information retrieval and thereby creating hurdles on efficient and fast filtering of information. To overcome the above problem, and to help the online readers to handle huge volume of data, a new method of information filtering has to be created. In this paper we present a new hybrid filtering approach which is the combination of collaborative and content based filtering. Collaborative Filtering is the method of matching people who have similar interest, which is done by collecting desirous information from many users using Pearson Correlation Coefficient. Content Based Filtering uses Cosine Similarity for analyzing the content of the user profile and the past browsing history maintained for each user. We present design of Adaptive Web Service Selection Framework that makes use of automated approach that systematically integrates all available training information such as past user, user similarity ratings as well content of user-paper rating, paper-user rating and user browsing history. The vital solution of our approach is simple relative frequency percentage bar to illustrate the combined approach which creates the most accurate web services recommendation to the individual user. The system is able to absorb the result of both filtering technique of information theories and it can be adopted for better and efficient recommendation. The results of this system reveals the fact that is able to dynamically offer recommended services based on end user’s interest.

Keywords— Collaborative filtering; Cosine similarity; content based filtering; Pearson correlation coefficient; E-commerce.

1. INTRODUCTION

In the present scenario of information oriented world, huge contents of data on various subjects are readily available in the internet. It is highly a complex task to extract useful data from the internet based on user’s choice within a very short period of time. To overcome this problem, a web site can be personalized which can help them to extract or download the preferred data as per their requirements. The recommender system will be a useful site made exclusively with the user’s preferences in mind. To handle the increase of day to day innovations of information technology and the changing way of human beings life styles, the recommender system provide a unique service by serving as an agent to get essential user’s information from the internet in the fastest manner. So, in the present world this recommender system would be very useful system for web browsers. The commonly used search engine by users in the present scenario is a search engines which gather information after comparing the preferences of maximum users.

In this paper, we describe our approaches of content based filtering and collaborative filtering. It is a new method of filtering approach which is the sophisticated one in the sense that it is unifying both filtering together to provide customized recommendation to the users on their field of interest. In the content based recommendation on the basis of a user’s preference, we gather the information of both users and paper preference. In collaborating recommendation people often get best recommendations from someone with their resembling preferences. It explores uniqueness among similar people with similar enthusiasm and making recommendation on this basis.

The core area of this paper is related to domain of web services with paper recommendation system as such we present a list of research papers in the database. In this paper recommendation system, user selects his/her paper from the database and matching similar preferences from other users available in the database. To obtain a user rating of paper, preferences from different users is considered and the prediction is done based on the Pearson Correlation Coefficient. Along with this, the system also analyses the user browsing history, user profile and paper profile to correlate all together and make prediction on the basis of cosine similarity. Finally it absorbs the result of both approaches and create more accurate suggestion to a user. We design advanced web service selection framework where the system find out relative frequency percentage bar which display the highly recommended list of paper to user. We present search engine - PLING (perfect legitimate intelligent new generation) in which user can enter their paper name according to their desirous information and both of these filtering perform their function and at the end our approach will dynamically offer web services that fit to the end users interest.

The rest of the paper is organized as follows: Section II contains the brief information of the related works in the area
of various filtering methodologies done in the recent past. Section III describes the proposed system and algorithms adopted for implementing the system, with a system architecture. Section IV elaborates the results of the system implemented with the dataset. Finally, in the Section V conclusion part, we have the summary of the new filtering approach used in the system and the scope for future work.

II. RELATED WORKS

This section explains our work that is going based on web service choices. Recommendation term usually rising a problem when the user really wants to extract relevant content from internet. Emergence of mass content available in the internet have created problem in identifying the best recommendation services to user.

Usually recommendation systems are classified into three categories: collaborative filtering, content based filtering and hybrid filtering. Michael J.Pazzani implemented advanced framework in which they created profile of user interests and discuss how each type of information is used individually and in the combined form for the recommendation from multiple sources [2]. In Fab System, content analysis is employed to generate user profile from web page ratings [4]. They address the explanation interface for automated collaborative filtering system and explain how and why they should implement. Hellocker [5] tried to extract meaningful explanation from computational model that are more adhoc than rule based expert systems and to provide a usable interface to the explanation.

Mainly researcher concentrate their works on building recommendation system based on collaborative filtering [5]. Thereafter they elaborate filtering approach into content based and other techniques such as demographic, knowledge based) were implemented. Among these approaches content based and collaborative filtering approaches are most two famous recommendation methods. Then Justin Basilico and Hofmann, presented novel learning architecture for the problem of predicting user ratings [6] based on the idea of defining kernel functions over user-items pairs. They [7] introduce new approach of collaborative and content based filtering in that they select information based on semantic content and calculating user-user similarity based on integrated information extracted from user profile and user rating (Byeong Man Kim).

Unique recommendation algorithm via collaborative tags of Heung-Nam lam presented a new approach to provide enhanced recommendation quality and it tries to overcome some of the limitation of collaborative filtering [8]. Prem Melvik describes content boosted collaborative filtering in which they enhance existing user data and then provides personalized recommendation through collaborative filtering and it performs better than pure content naïve hybrid approach [10]. They present three-way aspect model that systematically collaborate classic collaborative and content based recommendation. Both of these simultaneously consider the similarities of user rating and semantic content of web service and validated by extensive experiment studies using 3,693 real world web services publically available from internet [9].

Resnick introduce first approach on recommendation using collaborative filtering [11]. In this author described about collaborative filters which helped the people to make choices based on the opinions of other people. Group Lens [11] is a system for collaborative filtering of Netnews; it helped the news reader clients to find articles and along with the predicted scores and it made it easy for users to rate article after they read.

Our approach adapt the new model which builds paper recommendation system, which is sophisticated and adaptive one where system combines both of content based and collaborative filtering methods considered based on recommended research papers and it helps the user to finding useful information, display highly recommended research paper. When user searches some items in background both approaches works together. Content based method will take user preferences and collaborative based recommendation takes rating of different users given to each papers and make prediction on that basis.

III. PROPOSED SYSTEM AND ALGORITHMS

Nowadays recommendation systems are correlated to a large scale of disciplines, though primary spotlight of the system was on the entertainment domain and online shopping sites. Presently their usage is being enlarged to amusing network sites-learning and travel domain. The filtering is being used by distant researcher and clarification in distant ways. The target area of this paper related to paper recommendation system. It helps users in managing their reading list by learning preferences. Here hybrid approach enforced together, where collaborative filtering which gives a list of recommendation based on user rating in a database and in content, we find out web services that provides paper which user have preferred in the user profile and item profile from database and combine both approach together recommend best paper to user which tells user specifically what should he read next.

In this section, the core research of this system on web service recommendation is considered for selecting research papers of user’s choice. In the present era many multi-disciplinary research is gaining more popularity. The researchers find it difficult to search the related works published by different service providers. This system is implemented to list out best related papers based on the user choices and similar user’s service ratings. The system briefly introduce typical recommendation techniques which adopts the following filtering methods: Collaborative Filtering, Content based Filtering and Hybrid Filtering technique. The above system is implemented which takes into consideration the user preferences, paper specifications and the preferences of other users.
The above figure 1 shows the different components involved in our recommendation system. The user selects favorite recommendation link component then it goes to personalized user interface thereafter users can make request to search engine. User’s interest and preferences are stored in the repository and those information is retrieved when required and they are used in the collaborative and content based recommendation. They calculate both similarity and correlation between user’s preferences and paper profiles thereafter combine both the results together and dynamically offer papers which fit the end user’s interest.

A. Collaborative Filtering

Collaborative filtering is a method by collecting data based on similar choice/interest of many users. The papers recommended to a user are those preferred by similar user, which focus on the relationship between users and paper. Similarity of paper is determined by the similarity of the ratings of those papers by the users who have rated both papers. Database can be maintained on the basis of interest of many users on their ratings on various papers. Measuring interest on similarity of user’s can be rated by using Pearson correlation coefficient algorithm.

1) Pearson Correlation Coefficient

Similarity between two users can be accurately calculated with the Pearson correlation. This algorithm measures the linear dependence between two variables or users as a function of their attribute/qualities. It is considered to be an optimized and it quite famous algorithm for collaborative filtering. A popular similarity measure in user-based collaborative filtering - Pearson correlations as follows:

\[ \text{Similarity} = \frac{\sum_{p \in P} (r_{ap} - \bar{r}_a)(r_{bp} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{ap} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{bp} - \bar{r}_b)^2}} \]  

(1)

- where \(a\) and \(b\) are users, \(P\) is the set of papers.
- \(r_{ap}\) and \(r_{bp}\) are the rating given for the paper \(p\) by user \(a\) and \(b\) respectively.
- Similarity value ranges between 1 and -1.

Algorithm for Collaborative filtering

\[
\begin{align*}
\text{Paper} & \left( \right) \\
\text{Common users} & \left( \right) \\
D & \rightarrow \text{Dataset} \\
\text{Collaborative filtering}\left( \text{input}\right) \\
\end{align*}
\]

For all users in \(D\)
- Common users = users who read input
End For
For all Common users
- Paper = distinct paper read by common users
End For
For all paper in Paper
- \(X = \) rating of book by common users
- \(Y = \) rating of input by common users
- \(\text{Pearson} = \) Pearson \((X, Y)\)
Return highest correlation paper
\[
\text{Pearson} (x, y) \text{ Correlation coefficient} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

(2)

Return highest correlation coefficient

Suppose Table 1 is our sample dataset which contains ratings given by each user to each paper.

<table>
<thead>
<tr>
<th></th>
<th>Paper 1</th>
<th>Paper 2</th>
<th>Paper 3</th>
<th>Paper 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>10</td>
<td>8</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>User 2</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>User 3</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>User 4</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

A software engine is newly implemented to check the similarity of users amongst their preference of research and it is named Pling search engine. The new recommendation engine is based on the basic principles of correlation coefficient.

The Pling engine is capable to show the similarity among various type of paper readers, based on their assessment of rating given by them for the papers they have read. Initially the name of paper read by users on the basis of their rating on each paper is compiled and entered in the program. Thereafter, the search engine will automatically calculate the preferences and rating on each paper, calculate the correlation coefficient and suggest the list of papers with higher correlation.

B. Content Based Filtering

Content-based filtering builds a recommendation on the basis of a user's preference. The focus on content based is to know about contents of both user profiles, paper profile based on the keywords. It also uses the past history with the user preferences specified. Firstly, we have to create a user profile and paper profile using content of shared attribute space.

For example, in our paper domain, we represent it with the keywords present the given paper and other papers (using binary code). For user profile, we can do the same thing based on the user’s preferences and the given paper profile information. Then the similarity of user and paper can be figured out using cosine [13] similarity.

Suppose Table 2 is our paper profile (0 preferred keyword,
Table 2 User-Paper specifications

<table>
<thead>
<tr>
<th>Paper genres</th>
<th>User Paper features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>User1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 1 0</td>
</tr>
</tbody>
</table>

The following Table 3 is our sample paper-paper profiles which includes the features/keywords of the preferred paper and the existing set of paper profiles in the database.

Table 3 Paper-Paper feature specifications

<table>
<thead>
<tr>
<th>Paper genres</th>
<th>Paper-Paper feature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paper1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 1 0</td>
</tr>
</tbody>
</table>

The cosines similarity between user1 and paper1 are calculated using the given formula:

\[
\text{CosineSim}(\text{User1, Paper1}) = \frac{\sum_{i=1}^{n} \text{User1}_i \times \text{Paper1}_i}{\sqrt{\sum_{i=1}^{n} \text{User1}_i^2 \times \sum_{i=1}^{n} \text{Paper1}_i^2}}
\]

where \( n \) is the number of features, \( \text{User1}_i \) is the preference given by the user for the \( i^{th} \) feature and \( \text{Paper1}_i \) is the specification for the \( i^{th} \) feature of that paper.

Algorithm for Cosine Similarity

\[
\text{Similarity} \{U \}
\]

\[
\text{U} \rightarrow \text{user profile}
\]

\[
\text{D} \rightarrow \text{paper profile}
\]

\[
\text{u} \rightarrow \text{Dataset of current user from U}
\]

\[
\text{cosine similarity} \ (u, D)
\]

\[
\text{For all paper in D}
\]

\[
\text{Similarity}(u, D) = \frac{\text{dot-product}(u, D)}{\text{u} \ast \text{D}}
\]

\[
\text{Return highest similarity paper}
\]

Sample calculation is shown below:

\[
\text{Similarity}(\text{User1, Paper1}) = \frac{3 \times 1 + 1 \times 1 + 1 \times 1 + 1 \times 0}{\sqrt{3^2 + 1^2 + 1^2 + 1^2 + 1^2}} = 0.416
\]

As and when the users name is entered into the system, the Pling search engine would show User Id, if it exists, otherwise a User Id will create automatically by search engine for the user. Thereafter entire preferences of that user will compare with each keyword of each paper in paper profile file using cosine similarity algorithm. Finally, it displays highest similar papers to users.

'Pling search engine' keeps all the records of different users, even if, the user read one or two paper among complete column of papers, and calculate keyword preference accordingly.

C. Hybrid Filtering

Hybrid approaches that combine collaborative and content based filtering. It incorporates the result of both approaches that will create potential for a more accurate recommendation to user. The key integrates of our system is simple relative frequency percentage bar which will help the user to find out better and efficient result according end-user by user similarity and recommending similar paper to user. We design search engine - Pling recommendation system which works like similar to search engine but different from Google, Yahoo, etc. Our approach implements an enhanced method which help the user to have quality browsing. The system list the best paper selected using jquery and suggests best paper by eliminating unwanted items.

IV. RESULTS

The system is tested with the dataset of user and paper profiles. When the user submits their choice for a particular paper is given, the adaptive checker system first retrieves the ratings dataset and performs the correlation based filtering. In figure 2, we show result of the Pearson Correlation Coefficient which provides the relation of paper but, it seems to be less efficient as it does not take users preferences.

![Fig2 Result of Pearson Coefficient](image)

After PCC, the similarity checking based on the content of the papers is examined which considers both users and paper references and specifications. The paper given is compared with other papers and the user’s preferences is also considered. Figure 3 shows the results of content similarity which takes into consideration the user profile, and paper profiles.
Whereas, figure 4 shows comparative results that reflects both similarity and correlation. Both are considered to have relevant significance, as correlation helps to select papers read by similar users and similarity helps to select papers which provide most appropriate one with maximum features expected by the user.

So, in order to incorporate the advantage of both the filtering methods the hybrid approach is calculated using relative frequency methods. The combination using relative frequency method shows tremendous best result and that result is shown figure 5. The illustrated combined result is much more efficient than other two methods when it is performed individually.

The results of our system helps to rank the papers based on the similarity scores of hybrid methods. The papers with the highest similarity scores and considered to be the nearest ones and they are listed as suggestions to the user.

V. CONCLUSION AND FUTURE WORK

As growth of available amount of information in internet increases daily, it is difficult to find out relevant information. We present a novel method to assist the users to find the useful web services which provides the relevant information. The system implements two different approaches and put them within a common framework. Both approaches have their own advantage and disadvantage but combination of this approach shows efficient and better result compared to individually approach. PLING using jQuery list out the relevant information to user it has ability to add user and user can register their preferences. The system is tested with the paper domain as it is easy to gather the details of research papers and it would the researchers to identify the research work done in topics of their interest in an easy and adaptable manner.

In the future, the filtering process can be extended to include the semantic web search, so that the user can specify any keyword which he is familiar and the system could perform search based on all other keywords equivalent to word given by the user. Different users may use different words to represent the real world entities, semantic of web ontology would help create more efficient filtering system.

References


